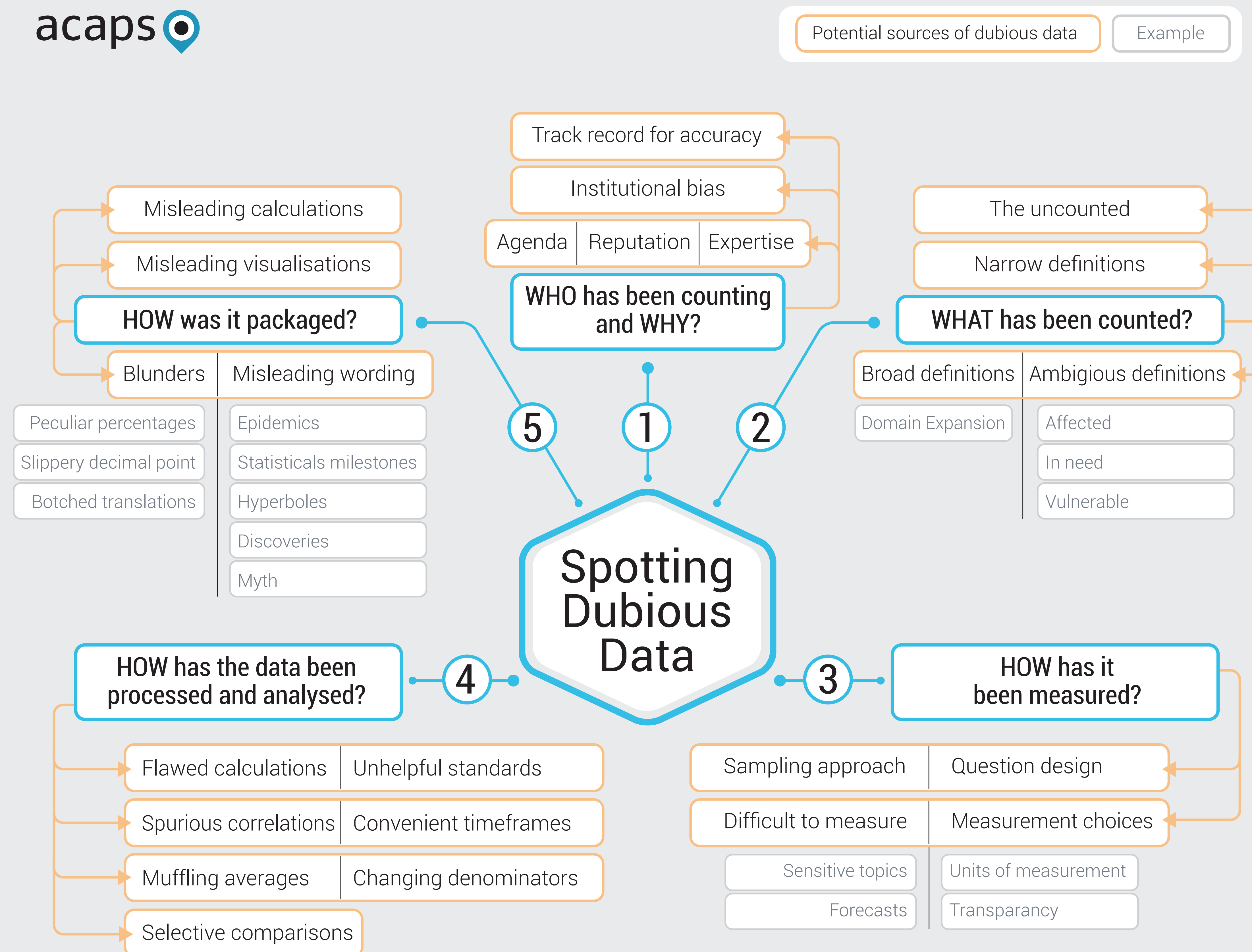




SPOTTING DUBIOUS DATA

Amid the volumes of information available on humanitarian crises, there are only few statistics worth remembering and using. Look out for the following potential sources of errors, scrutinize the data and spot the difference between solid stats and dubious data (Adapted from Joel Best).

SPOTTING DUBIOUS DATA



Adapted from Joel Best

SCRUTINISE THE DATA

1. WHO HAS BEEN COUNTING AND WHY ?

Example: Headline 14 September 2015, The Daily Mail: Two in every 100 Syrian migrants are IS fighters, according to the Lebanese Minister of Education. **Why is it dubious:** It is unlikely that the Lebanese Minister of Education has the expertise speak to the ratio of IS fighters to individuals fleeing Syria. It is likely that the Daily Mail, by some described as a 'sensationalist' newspaper, did not check this fact before publication.

Keep in mind:

- Why was the data collected? What is the agenda of the source? Could it be biased?
- What is the expertise of those who have collected, reproduced and disseminated the data?
- Are they sufficiently knowledgeable to research the matter?
- Is there a strong track record of producing accurate information?

2. WHAT HAS BEEN COUNTED ?

Example: In 2015, Colombia has the second highest number of IDPs in the world (6 million) after Syria (6.5 million). **Why is it dubious:** The concept of an IDP in Colombia is very broadly defined - displacement figures for Colombia commonly count all people who were internally displaced since the 1990s.

Keep in mind:

- Look out for concepts that are widely used within the humanitarian community, but lack a common definition such as affected, in need, vulnerable, household, urban.
- Consider whether the concepts used could have been defined too narrowly or too broadly. Has something been excluded?
- Have definitions remained the same at the different points in time? Has there been domain expansion? (Definitions that have been broadened over time?).

3. HOW WAS IT COUNTED ?

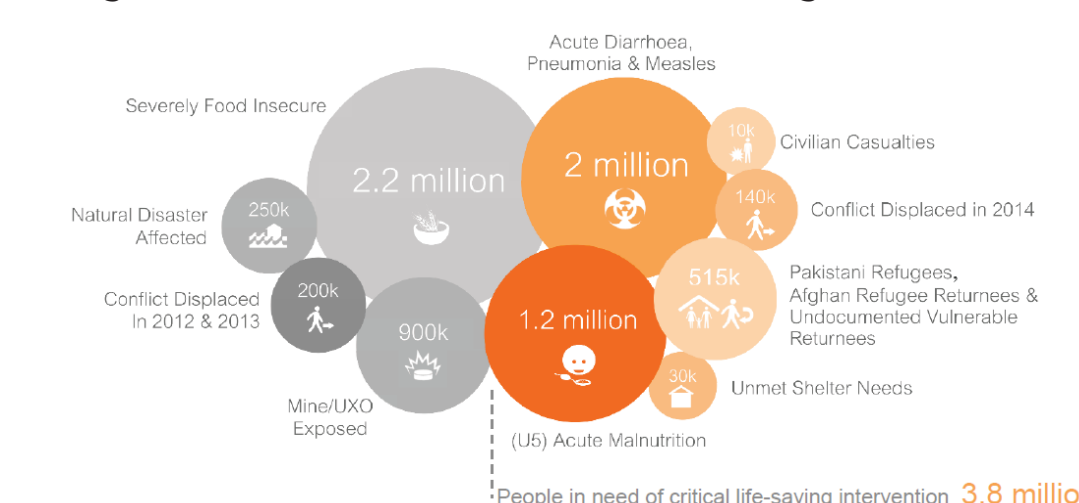
Example: 6.5 million people have been internally displaced in Syria as of October 2015. **Why is it dubious:** Data gathering in Syria is severely hampered by the active conflict and lack of access to parts of the country. (IDMC 07/2015) Statistics regarding the Syria conflict are therefore broad guesstimates, computed in a politically charged context.

Keep in mind:

- Does the data consist of numbers that seem hard to produce—how could anyone calculate that? Closely scrutinise information on sensitive topics, such as SGBV or informal activities.
- Numbers presented without sufficient information about measurement choices or assessment tools?
- Unusual units of analysis (e.g. extended families instead of households) that might affect the resulting statistic?
- Criticisms of measurement choices by others
- Particular caution is required when reviewing forecasts or estimates about future trends

4. HOW WAS IT PROCESSED AND ANALYSED ?

Example: 7.4 million people are in need in Afghanistan. in 2014, 7.4 million people are considered in need in Afghanistan, as shown in the following bubble chart.



Why is it dubious: Double-counting the number of people in need is common and this example is illustrative of the underlying thinking-error. The number of people in need per sector has been combined to total 7.4 million. However, the units of analysis are not mutually exclusive categories - some people who are severely food insecure, will have been affected by natural disasters too, etc.

crisis situation is often compared to the reference standards of those that organizations want to provide funding. The United States is one of the countries with the highest per capita water use in the world and is therefore not an appropriate comparison group. Sphere standards put total basic water needs per person per day at 7.5 to 15 liters a day.

Keep in mind:

- Could the calculations be flawed?
- Are there any misleading comparisons, timeframes, comparison groups or standards used?
- Are there any stated relationships between two variables (look out for reports that claim to identify the key cause of complex problems, it is impossible determine causality through experimental design)
- Calculations that highlight or muffle outliers?

Example: Water shortages for refugees in camps in Jordan have reached emergency levels; the supply is as low as 30 liters per person per day — one-tenth of what the average American uses. **Why is it dubious:** A

5. HOW WAS IT PACKAGED ?

Example: Of the more than 80 million people estimated to have been in need of humanitarian assistance in 2014, over 75% were women and children. **Why is it dubious:** 75% of all people in high fertility countries are women and children — it is unclear how this was calculated and it is most likely only included for shock purposes.

Keep in mind:

- Dramatic statements that take the form of statistical claims, such as hyperboles, 'the best', 'the most', 'myth', 'new discovery'?
- Unhelpful denominators (x per hour) used for shock purpose?
- Have results been misinterpreted? Are visual representation accurate or misleading?
- Blunders (numbers that seem surprisingly large or small)?
- Are the figures in line with what I know and expect or surprisingly different? Have decimal points been misplaced?

Example: Before the outbreak of violence in Burundi following mass-protest, under 5 global acute malnutrition rates were already at 41%. **Why is it dubious:** Global Acute Malnutrition (GAM) rates above 15% are considered critical, the most severe level of the WHO scale. One of the highest levels of GAM recently recorded was in South Sudan, at 22.7% (Generation Nutrition 2014)